IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Deb K. Chatterjee

Appl. No.: 09/558,421

Filed: April 26, 2000

For: **Mutant DNA Polymerases and Uses Thereof**

Art Unit: 1652

Examiner: Rao, M.

Atty Docket: 0942.3600003/RWE/BJD

#9
A.9.2
2/27/02

Declaration of Harini Shandilya

Commissioner for Patents
Washington, DC 20231

Sir:

I, Harini Shandilya, DO HEREBY DECLARE AND SAY:

THAT, I am presently employed by Life Technologies, Inc. (hereinafter "Life Technologies," and now Invitrogen Corporation)[1], the assignee of the above-captioned application as a Scientist. I have an M.S. degree in chemistry. I have been employed by Life Technologies since June 1, 1988.

THAT, one of my supervisors at Life Technologies has been Deb K. Chatterjee. During the time period the experiments described in this Declaration took place, I worked in close physical proximity to Deb Chatterjee. During this time period, I had at least a generalized knowledge of Deb Chatterjee's research projects.

THAT, at times Deb Chatterjee would ask me to witness his laboratory notebooks. When he asked me to witness laboratory notebook pages, I would carefully read and understand the content of those pages, then sign and date the pages at the lower lefthand corner. Deb

[1]Life Technologies, Inc. merged with Invitrogen Corporation on September 12, 2000, with Invitrogen Corporation being the surviving entity.

Chatterjee's notebook would often contain original data, e.g., photographs of gels. When original data was included in notebook pages that I witnessed, I would review the data together with the remaining content of the pages.

THAT, prior to October 17, 1994, 1 was aware that Deb Chatterjee had worked on a research project relating to mutant DNA polymerases. I was aware of this research project from various sources. These sources included informal discussions in the laboratory with Deb Chatterjee, discussions I heard between Deb Chatterjee and others, and my review of Deb Chatterjee's notebook, including original data included in that notebook.

Attached to this Declaration are several Exhibits containing laboratory notebook pages recorded by Deb Chatterjee. During the time period these experiments were conducted, as indicated on the notebook pages, I saw Deb Chatterjee performing molecular biology experiments. The nature of these experiments is consistent with the notebook entries and the cloning of a mutant DNA polymerase gene.

Attached to this Declaration as Exhibit A is a copy of a laboratory notebook page recorded by Deb Chatterjee. At the bottom lefthand corner of this notebook page I recognize my signature, indicating that I witnessed and understood this page on the date indicated. While the date is masked on the attached Exhibit, the date on the original is before October 17, 1994. Therefore, I conclude that this notebook page was in existence, and its content was communicated to me, before October 17, 1994.

The content of Exhibit A relates to a research plan to make *Taq* DNA polymerase mutants. I note that this page indicates that a distinctive difference between T7 polymerase, which utilizes dNTP and ddNTP equally well, and *Taq* polymerase in the presence of a Tyr in

lieu of a Phe at a location in the O-helix corresponding to Phe_{762} of *E. coli* polymerase I (Klenow fragment). This page also indicates to me that *Taq* DNA polymerase mutants would be made based on these differences, in order to construct a *Taq* polymerase that is more T7-like, e.g., would use dNTPs/ddNTPs equally well. I conclude that this page teaches a plan for making a *Taq* DNA polymerase mutant containing Tyr in lieu of Phe at a position corresponding to the Phe_{762} of *E. coli* polymerase I (Klenow fragment). Thus, the concept of this mutation had been communicated to me prior to October 17, 1994.

Attached to this Declaration as Exhibit B is a copy of two laboratory notebook pages recorded by Deb Chatterjee. At the bottom lefthand corner of the notebook page I recognize my signature, indicating that I witnessed and understood the page on the date indicated. While the date is masked on the attached Exhibit, the date on the original beside my signature is before October 17, 1994. Therefore, I conclude that these notebook pages were in existence, and their content was communicated to me before October 17, 1994.

The content of Exhibit B relates, among other things, to *Taq* mutant F667Y. The nomenclature "F667Y" refers to a substitution of Tyr (Y) for Phe (F) at position 667 of the *Taq* polymerase. This substitution is the same substitution referred to in Exhibit A. However, in this Exhibit the exact amino acid residue of *Taq* polymerase, instead of the corresponding *E. coli* polymerase I (Klenow fragment) residue, is explicitly identified. Thus, the concept of a specific mutation, $Phe_{667} \rightarrow Tyr_{667}$ of *Taq* polymerase, was conveyed to me prior to October 17, 1994.

Exhibit B also describes a mutagenesis experiment that was actually performed. In this experiment, oligonucleotide 2680 was annealed to a single-stranded DNA containing the wild-type *Taq* polymerase gene at 70°C for 2 minutes followed by cooling to about 35°C over

30-40 minutes. The resulting hybridized molecule was incubated with T7 polymerase and T4 ligase to synthesize and ligate the strand complementary to the single-stranded DNA using oligonucleotide 2680 as a primer, and then used to transform bacteria. DNA was then isolated from the transformed bacteria. This DNA was assayed for the presence of an additional *Ase*I restriction site derived from the oligonucleotide sequence. The presence of this additional restriction site would indicate incorporation of the oligonucleotide sequence, containing the F667Y mutation, into the mutagenized clone. Page 2 of Exhibit B contains original data showing the results of an assay of various mutagenized clones for the presence of the additional 590 bp *Ase*I restriction fragment, which would indicate the presence of the *Ase*I restriction site derived from the oligonucleotide sequence.

This page of Exhibit B is signed by me as having been witnessed and understood by me prior to October 17, 1994. Therefore, I would have reviewed this original data prior to October 17, 1994. Thus, experiments designed to construct this mutation and test it for the incorporation of an additional *Ase*I restriction site were performed prior to October 17, 1994.

Attached to this Declaration as Exhibit C is a page from Deb Chatterjee's notebook depicting an assay of additional clones for the presence of the additional *Ase*I restriction site. I have independently reviewed the original data attached to this page of Deb Chatterjee's notebook. Based on this independent review, I conclude that clone 8 contains an additional *Ase*I restriction site.

THAT, I have reviewed the oligonucleotide sequence identified as No.2680 on the Request for Synthesis of Oligonucleotides, which is attached to this Declaration as Exhibit D. I have also reviewed the publication of Lawyer et al., J Biol. Chem. 264(11):6427-6437 (1989),

which is attached to this Declaration as Exhibit E. In Figure 2 of Lawyer *et al.*, the DNA sequence of the *Taq* polymerase gene is disclosed. I have compared the DNA sequence in Figure 2 of the Lawyer article surrounding codon 667 with the sequence of oligonucleotide 2680 of Exhibit D to this Declaration. Based on this comparison, I have determined that the two sequences are complementary with two exceptions: (1) oligonucleotide 2680 contains a sequence complementary to a Tyr codon instead of the Phe codon in the sequence of Figure 2; and (2) oligonucleotide 2680 contains a sequence corresponding to an *Ase*I restriction site not found in the sequence of Figure 2.

THAT, I hereby declare that all statements made herein of my own knowledge are true; and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code; and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

2/15/02
Date

Harini Shandilya

Annealing:

	Kinased Oligo 2680	Kinased oligo 2723	Control
	1µl	1µl	—
Ann. Buffer	1µl	1µl	1µl
U-ss DNA (gene cleaved)	1.5µl	1.6µl	1.5µl
H_2O	6.5µl	6.5µl	7.5µl.

Anneal at 70°c for 2min → let cool to ~35°c over 30-40 min.

Synthesis:

Add to tubes. 1µl 10x Syn. Buffer
0.3µl T7 DNA Pol (USB)
1.0µl T4 ligase (LTI)

5' on ice, 5' at room temp — 45 min at 37°.

Add 90µl TE (10mM Tris pH8.0, 0.1mM EDTA).

Transfect DH5αF'IQ:

Transform DH5αF'IQ 100µl with 1.5µl ligation/synthesis mix.
30 min on ice
42°c for 90 sec. → Split 10% and 90% for 2680 and 2723
Keep 100% for the control.

Add 60µl lawn cells + 3ml soft agar
- let the agar solidified
→ Incubate plates at 37°.

To Page No.

# of Plaques in the plates

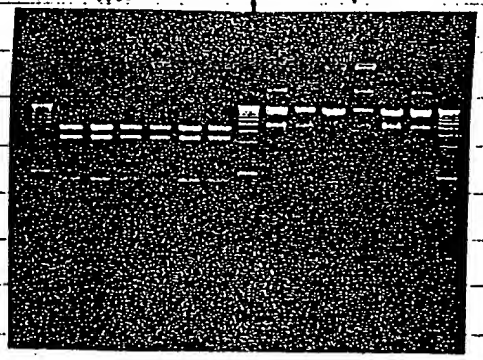
2680 (F667Y)	10%	90%	100%
	285	TMTC	—
2723 (Y671S)	250	TMTC	—
Control			150

Test 5 plaques from each, 50µl Lawn cells + Agar/Plaque Stab. Incubate at 37°C for 6 hrs!

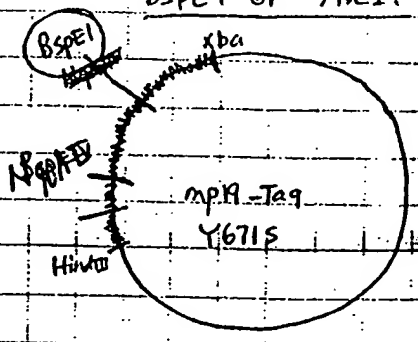
Standard miniprep - Take DNA in 120µl TE

µl was (Digested) F667Y with AseI NEB React3

7µl (") Y671S with BspEI NEB React3.



* Control RF cut with BspEI or AseI.



Y: AseI cut : expected ~590 bp fragment from one of the two larger fragment : Thus, are correct clone — Recheck few more clones.

s: BspEI cut : Expected a linear frag. #4 is the correct clone. #3 has an band (save both).

To Page I

essed & Understood by me,	Date	Invented by	Date
		Recorded by:	

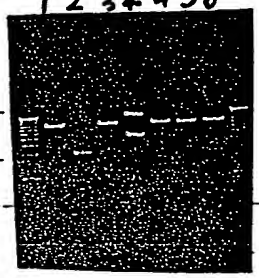
Exhibit C
Appl. No. 09/558,421

10% plating ≥ 500 colonies
90% plating TNTC.

Test 6 colonies.

* Control pTTQ18-Taq parent cut with BspEI - it does not have any BspEI.

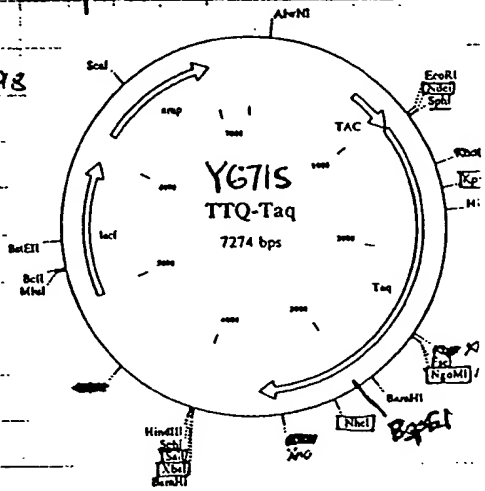
1 2 3 * 4 5 6



BspEI cut

#1, 3, 4, 5 & 6 are correct as they have the BspEI site (that was created)! &

The plasmids were pooled.

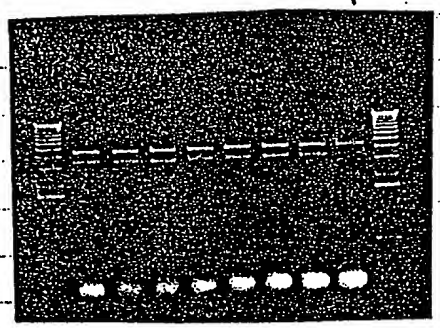


Save #1 & #3 as

Grow #1 & #3 for enzyme activity.

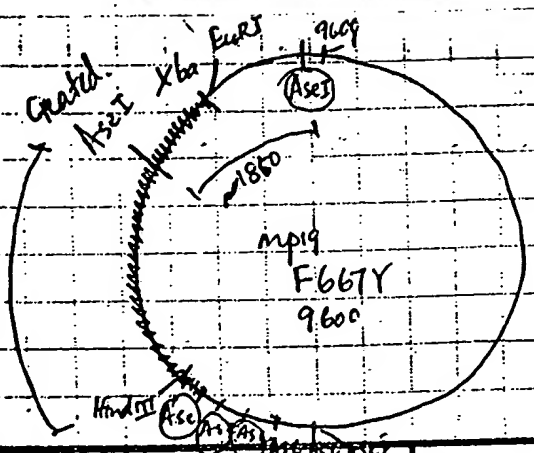
F667Y: Test 8 more plaques.

1 → 8



AseI cut

* #8 seems to be mixture of wild type and the mutant. The mutant should produce two bands ~1490 bp and ~1850 bp from the second largest band.



Thus #8 was saved. The fragment will be cloned as NgoAIV + XbaI and cloned into pTTQTaq. Then, look for an extra AseI site.

To Page No.__

ssed & Understood by me,	Date	Invented by	Date
		Recorded by	

Oligo# _____

REQUEST FOR SYNTHESIS OF OLIGONUCLEOTIDES

① #2679 ✓
② #2680 ✓

Date: _____

Name: Deb K. Chatterjee

Department: GENETIC ENGINEERING

Account: 0530 Cost: $180

Micromole Scale: 0.2 µ

① GAC | GGG | GAT | GTT | CGC | GAG | GTT | GGG | ATC #2679 (24-mer)
Nru I

② GTA | GAG | GAC | CCC | GTA | ATT | AAT | GGT | CTT | GGC | CGC
Acc I
#2680 (33-mer)

Brief Description Of Use:

#1 Q582A mutation of Taq

#2 F667Y mutation of Taq.

APPROVAL:

D Chatterjee
for DIRECTOR OR VICE PRESIDENT

FL
DATE RECEIVED

FL
DATE SYNTHESIS COMPLETED

THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 264, No. 11, Issue of April 15, pp. 6427-6437, 1989

Isolation, Characterization, and Expression in *Escherichia coli* of the DNA Polymerase Gene from *Thermus aquaticus**

(Received for publication, August 31, 1988)

Francis C. Lawyer‡, Susanne Stoffel‡, Randall K. Saiki§, Kenneth Myambo‡, Robert Drummond¶, and David H. Gelfand‡‖

From the Departments of ‡Microbial Genetics, §Human Genetics, and ¶Protein Chemistry, Research Division, Cetus Corporation, Emeryville, California 94608

The thermostable properties of the DNA polymerase activity from *Thermus aquaticus* (*Taq*) have contributed greatly to the yield, specificity, automation, and utility of the polymerase chain reaction method for amplifying DNA. We report the cloning and expression of *Taq* DNA polymerase in *Escherichia coli*. From a λgt11:*Taq* library we identified a *Taq* DNA fragment encoding an epitope of *Taq* DNA polymerase via antibody probing. The fusion protein from the λgt11:*Taq* candidate selected an antibody from an anti-*Taq* polymerase polyclonal antiserum which reacted with *Taq* polymerase on Western blots. We used the λgt11 clone to identify *Taq* polymerase clones from a λCh35:*Taq* library.

The complete *Taq* DNA polymerase gene has 2499 base pairs. From the predicted 832-amino acid sequence of the *Taq* DNA polymerase gene, *Taq* DNA polymerase has significant similarity to *E. coli* DNA polymerase I. We subcloned and expressed appropriate portions of the insert from a λCh35 library candidate to yield thermostable, active, truncated, or full-length forms of the protein in *E. coli* under control of the *lac* promoter.

Taq DNA polymerase (Taq Pol I)[1] isolated from *Thermus aquaticus* has been shown to be highly useful in the polymerase chain reaction (PCR) method (1, 2) of amplifying DNA fragments (3). The high temperature optimum activity, 75 °C, affords unique advantages when comparing Taq Pol I to *Escherichia coli* DNA polymerase I. High specificity of primer binding at the elevated temperature gives a higher yield of the desired product with less nonspecific amplification product. Also, *E. coli* DNA polymerase I is inactivated at 93–95 °C, the temperature range required to denature the duplex DNA product. Since Taq Pol I is stable at 93–95 °C, one can add Taq Pol I only at the beginning of the PCR reaction rather than before each round of amplification.

A 62–63-kDa Taq Pol I has been purified from *T. aquaticus*, but growing the organism is more difficult than *E. coli* and polymerase yields are low (4, 5). We have developed an alternative purification protocol[2] yielding a 94-kDa enzyme with 10–20 times higher specific activity than that previously reported. While the activity yield is quite high (40–60%), the initial expression level of *Taq* DNA polymerase in the native host is quite low (0.01–0.02% of total protein). Therefore, we sought to clone the Taq Pol I gene and express the gene in *E. coli*. In addition, the availability of the enzyme and the DNA sequence of the *Taq* DNA polymerase gene will facilitate the study of structure/function relationships and permit detailed comparisons with mesophilic DNA polymerases.

MATERIALS AND METHODS[3]

RESULTS

λgt11 Libraries—The construction of three λgt11:*Taq* libraries is described under "Materials and Methods," in the Miniprint. To maximize the probability of recovering a Taq Pol I epitope, three separate *Alu*I libraries were prepared. We ligated 8-mer, 10-mer, and 12-mer *Eco*RI linkers to the *Taq* *Alu*I DNA fragments to ensure that each *Alu*I fragment would be in-frame with respect to β-galactosidase in one of the libraries. Upon screening with primary antibody from Taq Pol I-immunized rabbits and plaque purification, we identified seven positive plaques from the 12-mer library, four positive plaques from the 10-mer library, and no positive plaques from the 8-mer library. The *Eco*RI inserts fell into four size classes: two of the seven phage isolated from the 12-mer library and two of the four phage isolated from the 10-mer library contained 115-bp inserts, five clones from the 12-mer library had inserts of 175 bp (one of these also had a second apparently unrelated *Eco*RI fragment of 185 bp), one clone from the 10-mer library had a 125-bp insert, and one clone from the 10-mer library had a 160-bp insert. Upon antibody screening each of the phage reacted with immune serum but did not react with preimmune serum. ^{32}P-labeled probes were prepared by PCR amplification (3) of one clone each from the 115-, 175-, and 125-bp size classes. The 115-bp probe hybridized with all the candidates containing 115-bp inserts and no others. Similarly, the 175-bp probe hybridized with candidates containing 175-bp inserts, and the 125-bp probe hybridized

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04639.

‖ To whom correspondence and reprint requests should be addressed: 1400 53rd St., Emeryville, CA 94608.

[1] The abbreviations used are: Taq Pol I, DNA polymerase isolated from *T. aquaticus*; kb, kilobase(s); bp, base pairs; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; dNTP, deoxyribonucleotide triphosphate; kDa, kilodalton; *X*-Gal, 5-bromo-4-chloro-3-indolyl-β-D-galactoside; IPTG, isopropyl-1-thio-β-D-galactopyranoside; PBS, phosphate-buffered saline; TMB, 3,3′,5,5′-tetramethyl benzidine; PCR, polymerase chain reaction; Pol I, DNA polymerase I.

[2] D. Gelfand and S. Stoffel, manuscript in preparation.

[3] Portions of this paper (including "Materials and Methods," Table V, and Fig. 8) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

with only the candidate containing that insert. Subsequent DNA sequencing of two 115-bp *Eco*RI inserts, one each from the 12-mer and 10-mer libraries, confirmed that they were identical sequences. DNA sequence analysis of *Taq* and flanking *lacZ* DNA for the candidate from the 12-mer library indicated the presence of one *Eco*RI linker at its 5′ *lacZ* junction. DNA sequence analysis of the *Taq* and flanking *lacZ* DNA for the 115-bp candidate from the 10-mer library indicated the presence of three *Eco*RI linkers at the 5′ *lacZ* junction, which resulted in the same frame with respect to β-galactosidase as that of the 12-mer linker candidate. Thus, we picked DNA fragments encoding the same epitope from tw libraries.

Lysogens were made of all the candidates in strain Y1089 and were induced with isopropyl-1-thio-β-D-galactopyranoside (IPTG). Total proteins from crude lysates of induced cultures were run on SDS-PAGE gels, and Western blots were prepared by using the anti-Taq Pol I antibody for detection. All of the clones made IPTG-inducible, *lacZ*-fusion proteins which reacted with the anti-Taq Pol I antibody (data not shown).

One clone each from the 115-, 125-, 160-, and 175-bp insert size classes was chosen for epitope selection. This method uses crude extracts of candidate clones to select antibodies from a polyclonal antiserum. These affinity-selected antibodies were used to probe Western blots of Taq Pol I. The results are shown in Fig. 1. In two experiments candidate λgt11 1, the 115-bp insert candidate, was the only one of the four tested which successfully bound antibody that reacted with purified Taq Pol I and reacted uniquely with Taq Pol I in crude extracts. The other three candidates, which had been identified and purified with the anti-Taq Pol I antibody, failed to "fish" from that same polyclonal antibody an antibody that would react with Taq Pol I on a Western blot. A close inspection of the Western blot indicates a faint cross-reaction with 28–30-kDa proteins in total soluble *Thermus* crude extracts. The DNA sequences of these three candidates do not correspond to any part of the Taq Pol I DNA sequence (Fig. 2).

λCh35 Libraries—The 115-bp *Eco*RI fragment from clone λgt11 1 was subcloned into Genescribe Z vector pTZ19R to use as a probe in screening the λCh35:*Taq* library. Construction of the partial *Sau*3A digest library of *Taq* DNA in λCh35 and screening of the library are detailed under "Materials and Methods," in the Miniprint. The *in vitro* packaged library was plated initially on *E. coli* strain K802. That strain was chosen to avoid the possibility of degradation of *Taq* insert DNA by the *mcrA* or *mcrB* restriction systems (6). The amplified library was subsequently plated on *E. coli* strain MC1000.

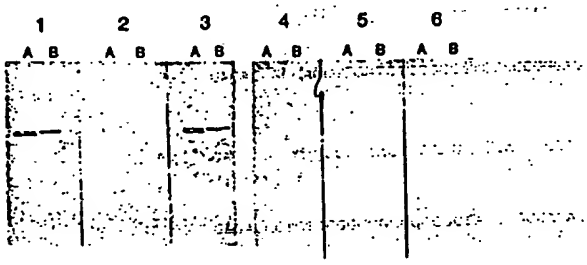


FIG. 1. Immunoblots with affinity-purified antibodies prepared by epitope selection. Epitope selection is described under "Materials and Methods." For each immunoblot, 3 units of purified Taq Pol I (partially proteolyzed) plus 10 μg of gelatin were loaded on *Lane A*, and 10 μg of *Taq* crude extract was loaded on *Lane B*. Antibodies used to probe immunoblots were: *1*, 1:10,000 dilution of the anti-Taq Pol I polyclonal antiserum; 2, anti-Taq Pol I antibody affinity purified with purified β-galactosidase (negative control); *3–6*, anti-Taq Pol I antibodies affinity purified with extracts of induced λgt11 clones 1, 3, 9, and 2–11, respectively.

Nine candidates were isolated and purified from the λCh35:*Taq* library. From restriction analysis of mini DNA preparations, none of the candidates proved to be identical, though they all shared some common restriction fragments. Upon Southern blotting, the pTZ19R:1 probe hybridized to a common 4.2-kb *Bam*HI fragment and a common 6.5-kb *Pst*I fragment in all the candidates, consistent with the hybridization seen in Southern blots of Taq genomic DNA (Fig. 3). For *Hin*dIII, the probe hybridized to fragments of different sizes, ranging in size from 5.6 to 10 kb. In addition, all nine candidates shared a common 4.5-kb *Hin*dIII fragment.

One candidate, designated ϕ4-2, had a probe-hybridizing *Hin*dIII fragment of approximately 8 kb which corresponded to the *Hin*dIII fragment that hybridized with probe 1 in the *Taq* genomic Southern (Fig. 3). We chose this candidate for further study and subcloned each of its four detectable *Hin*dIII fragments (A = 8 kb, B = 4.5 kb, C = 0.8 kb, and D = 0.5 kb) into vector BSM13⁺ in both orientations, transforming into host DG98. The two subclones of fragment A in both orientations, pFC82.35 and pFC82.2, were IPTG-induced and extracts were assayed for Taq Pol I activity (Table I). Subclone pFC82.35 had IPTG-inducible thermostable activity at a very low level, which was detectable because of the high sensitivity of the assay (<1 molecule/10 cell equivalents). In contrast, pFC82.2 had a significantly lower basal level of Taq Pol I activity which was attenuated in extracts of IPTG-grown cultures.

A restriction map of the A fragment was generated and is shown in Fig. 4. Southern analysis showed that the λgt11 1 probe hybridized at one end of the A fragment. Indeed, the DNA sequence of the *Alu*I genomic fragment isolated in λgt11 1 corresponds to nucleotides 619–720 in the Taq Pol I gene (Fig. 2). Further, the *Eco*RI-adapted *Alu*I site at the junction between *E. coli lacZ* and *Taq* in λgt11 1 corresponds to the *lac* promoter-proximal *Taq Hin*dIII site in pFC82.35.

Deletions in the A Fragment to Localize the Taq Pol Gene—Two different deletions were made in the A fragment in pFC82.35 to aid in localizing the gene. In pFC84, approximately 2.4 kb of the right end of the A fragment was deleted from the *Sph*I site (Fig. 4) rightward to the *Sph*I site in the vector polylinker. In pFC85, approximately 5.2 kb of the right end of the A fragment was deleted from the *Asp*718 site rightward (Fig. 4) to the *Asp*718 site in the vector polylinker, leaving 2.8 kb of *Taq* insert sequence. The activity of Taq Pol I was assayed in extracts of uninduced and IPTG-induced pFC84 and pFC85 in DG101. As can be seen in Table I, deleting 3′ sequences in the A fragment had a dramatic effect on the IPTG-inducible expression of Taq Pol I. In addition, while we were unable to detect Taq Pol I in Western blots of IPTG-induced pFC82.35/DG98, induced immunoreactive bands were clearly seen upon Western blotting of IPTG-induced pFC84/DG101 and pFC85/DG101 (Fig. 5). In the Western blots, induced pFC84/DG101 and pFC85/DG101 lanes revealed doublet immunoreactive bands that were approximately 65- and 63-kDa. These immunoreactive species were considerably smaller than full-length 94-kDa Taq Pol I. We determined that the doublet bands were not artifacts of the gel analysis because they were seen repeatedly in several experiments.

LacZα Fusions—To define further the locus of the Taq Pol I gene and to confirm the reading frame at different sites for use as guideposts during DNA sequence analysis, we constructed several fusions of the left end of the *Taq Hin*dIII A fragment to *lacZα* in the BSM13⁺ vector. These fusions are

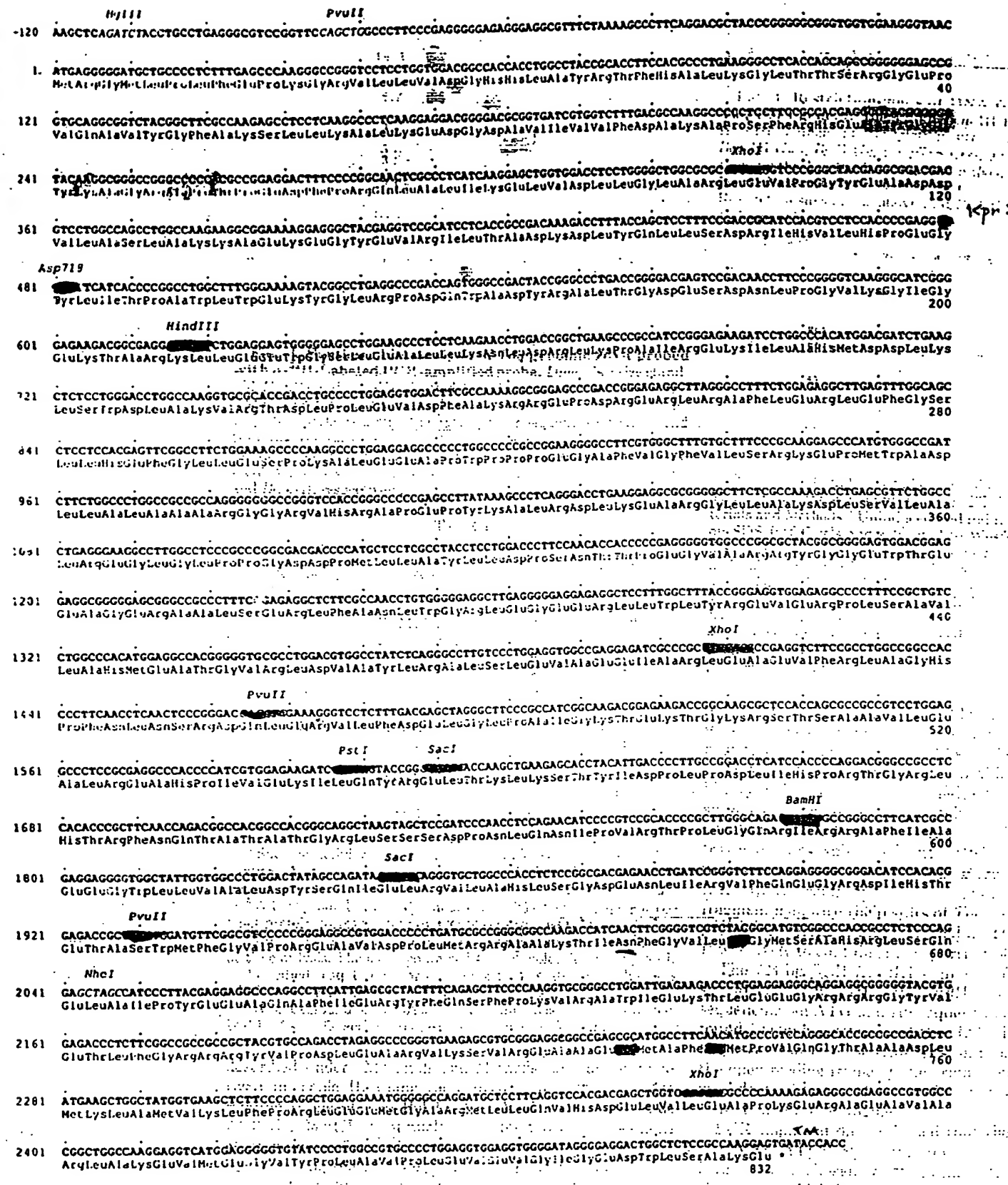


FIG. 2. DNA sequence and deduced amino acid sequence of the Taq Pol I gene. Nucleotides were numbered consecutively from the start of the gene. *Nucleotide numbers* are shown on the *left*. *Amino acid numbers* are shown on the *right*.

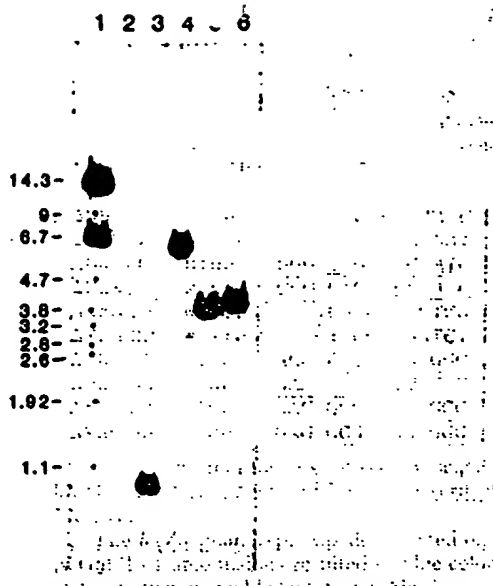


FIG. 3. Southern blot analysis of *Taq* genomic DNA probed with α-^{32}P-Labeled PCR-amplified probe. *Lane 1* is a size standard *Eco*RI- and *Bam*HI-digested λplac5 and *Msp*I-digested plasmid Lac5. DNA fragment sizes (in kilobases) are listed at *left*. The PCR-amplified probe contains the λgt11 primer sequences on either end (flanking the *Eco*RI site in *lacZ*) which are homologous to sequences in the 14,300 and 6,700 marker bands. *Lanes 2–6* are *Taq* genomic DNA digested with *Hin*dIII, *Hin*dIII and *Pst*I, *Pst*I, *Pst*I and *Bam*HI, and *Bam*HI, respectively.

TABLE I

Taq DNA polymerase activity in E. coli extracts

Experiment	Extract	IPTG	Specific activity[a]
I	BSM13⁺	±	<0.001[b]
	BSM13⁺w/Taq[c]	+	0.142
	BSM13⁺w/Taq[d]	+	0.136
	pFC82.35	−	0.248
		+	0.310
	pFC82.2	−	0.031
		+	0.002
II	BSM13⁺	+	0.003[e]
	pFC84	−	1.24
		+	29.7
	pFC85	−	0.87
		+	29.6
	pLSG1	−	4.4
		+	37.5

[a] Specific activity in units/mg total crude extract protein when assayed, as described under "Materials and Methods," on clarified, heat-treated extracts.

[b] A background of 0.004% input counts has been subtracted. Extract protein corresponding to 3×10^7 cells was assayed.

[c] Purified Taq DNA polymerase was added to a replicate cell pellet at time of lysis. The assay contained 4×10^7 molecules of Taq Pol I.

[d] Purified Taq Pol I, corresponding to 4×10^7 molecules, was admixed with the BSM13⁺ extract at time of assay.

[e] A background of 0.002% input counts has been subtracted. BSM13⁺ specific activity represents two times background.

described under "Materials and Methods" and are summarized in Table II. Using these fusions we determined the reading frame of Taq Pol I at the *Nhe*I site at nucleotide 2043, the *Bam*HI site at nucleotide 1780, and at four locations at or leftward of the *Xho*I site at nucleotide 1408.

Assembling the Full-length Taq Pol I Gene—As described above, the *Sph*I and *Asp*718 deletants, pFC84 and pFC85, produced thermostable polymerase activity upon induction. However, the size of the induced bands detected by anti-Taq Pol I antibody in Western blots was smaller than full-length

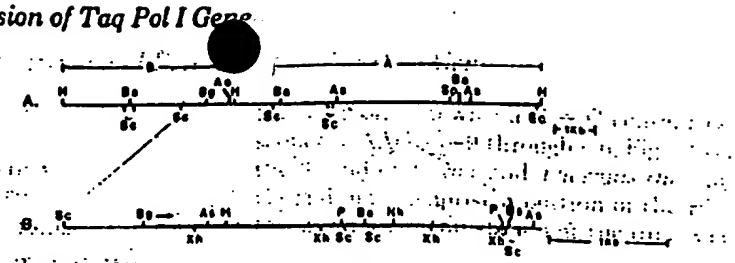


FIG. 4. Restriction maps of DNA fragments containing the Taq P I I gene. *A*, the 4.5-kb *Hin*dIII B fragment and the 8.0-kb *Hin*dIII A fragment. Restriction sites are: *Hin*dIII (*H*), *Sac*I (*Sc*), *Bam*HI (*Ba*), *Bgl*I (*Bg*), *Asp*718 (*As*), and *Sph*I (*Sp*). *B*, expansion showing the Taq Pol I coding region (*bold line*). *Arrow* (→) indicates N terminus of the gene. *Dotted line* (- - -) indicates λgt11-1 sequence. Restriction sites are as above and *Hst*EII (*Hs*), *Xho*I (*Xh*), *Pst*I (*P*), and *Nhe*I (*Nh*).

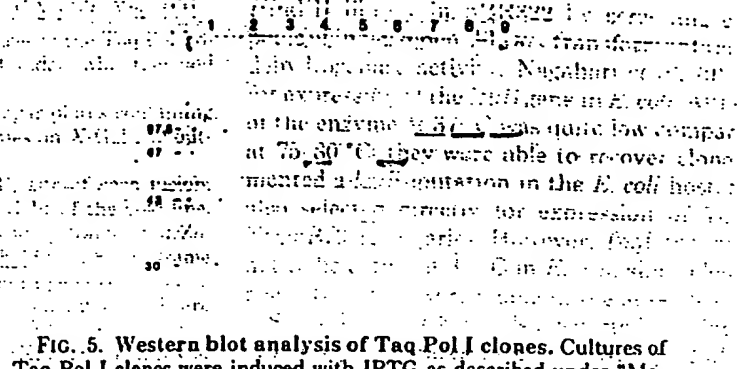


FIG. 5. Western blot analysis of Taq Pol I clones. Cultures of Taq Pol I clones were induced with IPTG as described under "Materials and Methods." Uninduced and induced samples were analyzed on SDS-PAGE gels and subjected to Western blot analysis, also as described under "Materials and Methods." *Lane 1*, Pharmacia low molecular weight marker. Molecular weights (in thousands) are shown at *left*. *Lane 2*, induced BSM13⁺ (33 μg) negative control. *Lanes 3* and *4*, uninduced (0.04 unit, 33 μg) and induced (1.0 unit, 33 μg) pFC85. *Lanes 5* and *6*, uninduced (0.03 unit, 33 μg) and induced (1.0 unit, 33 μg) pFC84. *Lanes 7* and *8*, uninduced (0.05 unit, 11 μg) and induced (0.4 unit, 11 μg) pLSG1. *Lane 9*, BSM13⁺ (33 μg) plus 0.4 unit of purified Taq Pol I.

Taq Pol I, i.e. approximately 65-kDa as opposed to full-length 94-kDa. Thus, we felt that the A fragment lacked the 5′ portion of the gene which would encode the N terminus.

Also mentioned earlier, all candidates from the λCh35 library which had been identified with the pTZ19R: I probe shared a common, approximately 4.5-kb *Hin*dIII fragment which did not hybridize to the probe. This fragment, the B fragment, was subcloned into BSM13⁺, yielding plasmid pFC83. The restriction map of the B fragment was determined (Fig. 4). By comparing those mapping results and the A fragment map with the results of *Taq* genomic Southern blots probed with probe 1 (Fig. 3) we deduced that *Hin*dIII fragment B was likely to contain the 5′ portion of the Taq Pol I gene.

The 724-bp *Bgl*II-*Hin*dIII segment of the B fragment was subcloned into *Bam*HI- and *Hin*dIII-digested BSM13⁺. Upon sequencing, an ATG and subsequent open reading frame was found 109 bp downstream of the *Bgl*II site. The open reading frame continued to the *Hin*dIII site. In addition, the phase of the open reading frame at the "right" end of the B fragment was identical to the phase of the open reading frame at the "left" end of the A fragment.

PCR amplification confirmed that the B and A fragments in pFC83 and pFC82.35 are contiguous in the *Taq* genome. Primers were chosen which flanked the presumed internal *Hin*dIII site: MK138 (Table V, in the Miniprint) on the left side of *Hin*dIII and FL25, a 20-mer complementary to nucleotides 622–641 of the Taq Pol I sequence, on the right side of *Hin*dIII. Upon amplification (3) of the λCh35 genomic phage

TABLE II
LacZα fusions

Fusion[a]	*LacZα* phenotype[b]	Fusion DNA sequence[c]								
		Taq				Polylinker				
ΔNhe 1	Blue			GAG	CTA	G	CG	AGC	TCG	
ΔBa 15	White			CAGAGGAT			CCC	CGG	GTA	
ΔBa 33	Blue	GGG	CAG	AGG	ATC		GAT	CCC	CGG	GTA
ΔBa 35	Blue	GGG	CAG	AGG	ATC		CCC	CGG	GTA	
ΔXh 28	White		TCGCCCGCCTCG				GTA	CCG	AGC	TCG
ΔXho 30	White		GTGGGCGGATCT				TA	CCG	AGC	TCG
ΔXho 32	Blue	AGG	CTT	GAG	GGG		GTA	CCG	AGC	TCG
ΔXho 53	Blue	GAA	GGC	CTT	GGC		GTA	CCG	AGC	TCG
ΔXho 54	Blue	GAG	GGG	GTG	GCC		CCG	AGC	TCG	AAT
ΔXh 59	Blue	GAG	GCG	CGG	GGG		GTA	CCG	AGC	TCG

[a] Construction of fusions between 5′ sequences of the Taq Pol I 8-kb HindIII A fragment and *lacZα* is described under "Materials and Methods."

[b] The *lacZα* phenotype was determined on agar plates containing *X*-Gal. In-frame fusions resulted in blue colonies on *X*-Gal and out-of-frame fusions yielded white colonies.

[c] The DNA sequence was determined at the site of each fusion. BSM13⁺ polylinker sequence is shown to the right of the bold line. Groupings of three nucleotides indicate the reading frame of *lacZα*. *Taq* DNA sequence is shown to the left of the bold line. For in-frame (blue) fusions, the deduced frame of the Taq Pol I gene is indicated. Restriction sites regenerated (*Nhe*I, *Bam*HI) or generated (*Cla*I) are indicated by italics. The Taq Pol I nucleotide coordinates (Fig. 2) at the fusion sites of the *Xho*I *lacZα* fusions are: ΔXho 28, 1411; ΔXho 30, 962; ΔXho 32, 1266; ΔXho 53, 1098; ΔXho 54, 1173; ΔXho 59, 1050.

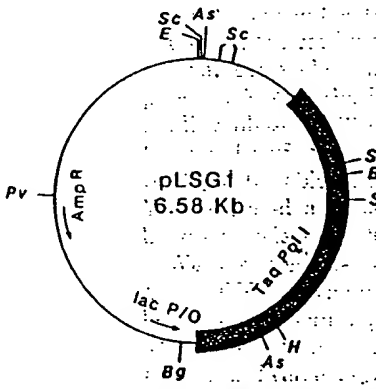


FIG. 6. Plasmid pLSG1. The 6.58-kb plasmid contains a 3.41-kb segment of *T. aquaticus* DNA in a derivative of the plasmid vector BSM13⁺. The *bold line* indicates the 2.5-kb Taq Pol I coding sequence. Expression of Taq Pol I is controlled by the *lac* promoter/operator. Construction of the plasmid is described under "Materials and Methods." Restriction sites are as in legend to Fig. 4 and *Eco*RI (*E*) and *Pvu*I (*Pv*).

ϕ4-2, we observed the predicted PCR product of 86 bp (data not shown), indicating that the B and A *Hin*dIII fragments are contiguous. A larger PCR product would have indicated that there was another *Hin*dIII fragment in the gene.

The assembly of the full-length Taq Pol I gene in plasmid pLSG1 (Fig. 6) is described under "Materials and Methods." Cultures of pLSG1 in DG101 produced IPTG-inducible thermostable polymerase activity at 37.5 units/mg protein (Table I). Western blots of IPTG-induced pLSG1/DG101 cultures revealed an immunoreactive band of appr ximately the same size as full-length Taq Pol I, 94-kDa (Fig. 5). Coomassie staining of IPTG-induced pLSG1/DG101 cultures failed to indicate the presence of a detectable induced band. The complete nucleotide sequence of the Taq Pol I gene and the deduced amino acid sequence are presented in Fig. 2. The DNA sequence of the Taq Pol I gene predicts an open reading frame of 2496 bp with a G + C content of 67.9%. The DNA sequence AAGG (−9 through −6, Fig. 2) is complementary to the 3′ end of *E. coli* and *Thermus thermophilus* 16 S rRNA (7) and may comprise a portion of the ribosome binding site for initiation of translation at the first ATG.

DISCUSSION

Several groups have reported the cloning and expression in *E. coli* of genes from thermophiles: malate dehydrogenase (*mdh*) from *Thermus flavus* (8), β-isopropylmalate dehydrogenase (*leuB*) from *Thermus thermophilus* (9), and the *Taq*I restriction-modification system from *Thermus aquaticus* (10). Iijima *et al.* (8) selected the *mdh* gene from a *T. flavus* partial *Hin*dIII library in pBR322 by screening crude extracts of pools of independent library transformants at 60 °C for malate dehydrogenase activity. Nagahari *et al.* (9) selected directly for expression of the *leuB* gene in *E. coli*. Although the activity of the enzyme at 37 °C was quite low compared to its activity at 75–80 °C, they were able to recover clones which complemented a *leuB* mutation in the *E. coli* host. Slatko *et al.* (10) also selected directly for expression of *Taq*I methylase in *Taq*:pBR322 libraries. However, *Taq*I endonuclease appeared not to be active at 37 °C in *E. coli*, since clones with only the restriction gene were viable in the absence of modification.

Several groups have also reported cloning and expression of DNA polymerases in *E. coli*. Kelley *et al.* (11) cloned the structural gene for DNA polymerase I (Pol I) from *E. coli* in λ bacteriophage. They observed polymerase activity in the transducing phage at a level of approximately 4% of total cell protein. However, they were unable to maintain a plasmid harboring the PolA⁺ gene, probably because overproduction of Pol I in *E. coli* is lethal to the cell. More recently, T4 DNA polymerase has been cloned and expressed in *E. coli* (12). In this case, it was necessary to clone the gene under control of inducible promoters such that constitutive expression of the gene would be minimal. Attempts to clone the gene under control of its own promoter in *E. coli* were unsuccessful, probably because of the detrimental effect the polymerase had on the cell. We did not know if Taq Pol I would be toxic to *E. coli* cells at 37 °C. While the *in vitro* specific activity of Taq Pol I at 37 °C is only a few percent of the specific activity at 75 °C, we could not predict if the DNA binding activity of the enzyme might interfere with normal cell function. To avoid potential problems related to direct expression of the gene in *E. coli* we chose to clone an epitope of the Taq Pol I gene by using λgt11 libraries and antibody selection. The epitope-expressing clone was subsequently used to select the entire Taq Pol I gene from a library in λCh35.

We were unable to detect a thermostable polymerase activity in cells infected (11) with any of the λCh35 clones, including ϕ4-2. The polymerase assay is extremely sensitive and can detect 1 molecule of polymerase per 10 cell equivalents. Upon subcloning of the 8-kb probe-hybridizing *Hin*dIII A fragment from ϕ4-2 into BSM13⁺ and IPTG induction of the subclone pFC82.35, a low level of thermostable polymerase activity was detected (Table I). Based on the activity of purified Taq Pol I when admixed with *E. coli* cells, this activity represents two to three molecules of Taq Pol I per cell equivalent. The gene was localized to one end of the 8.0-kb *Hin*dIII A fragment by using deletion analysis. Upon IPTG induction, pFC84, the *Sph*I deletion, and pFC85, the *Asp*718 deletion, yielded a 100-fold increase in Taq Pol I activity (Table I) compared to that of the full-length A fragment subclone, pFC82.35. This increase in activity allowed for r ady detection of the induced protein(s) on Western blot (Fig. 5). The A fragment induced proteins were truncated with an

apparent molecular mass f 63–65 k

Fusing the 5′ *Hin*dIII site in the A fragment with the *Hin*dIII site in BSM13⁺ causes the Taq Pol I gene to be out of frame with respect to β-galactosidase. The reading phase at the *Hin*dIII site in BSM13⁺ with respect to β-galactosidase is A AGC TT, a frame of "0" (13). The reading frame of Taq P l I at the *Hin*dIII site is AAG CTT ("plus 1"). The fusion gives rise to a minus 1 frame shift. In the β-galactosidase reading frame, there is a TGA stop codon at nucleotide 1478 of Taq Pol I. Downstream of this TGA there are several possibilities for restarts which could result in truncated forms of Taq Pol I: ATGs at nucleotides 1509 and 1752 and GTGs at nucleotides 1547, 1569, 1722, and 1731. In fact, we see a doublet in induced lanes of both pFC84 and pFC85 on Western blots (Fig. 5) indicating at least two reinitiation sites. All but one of the likely sites, the ATG at nucleotide 1509, would probably require a ribosome binding site for reinitiation. There are reasonable ribosome binding sites for the GTG at nucleotide 1722 and for the ATG at nucleotide 1752. Translation initiating at these sites would yield proteins of 59 and 58 kDa, respectively. However, the apparent molecular masses f the doublet bands seen on Western blots of pFC84 and pFC85 are approximately 65 and 63 kDa, based on comparison of the mobilities of the doublet bands with the molecular weight size standards. Whether the result of reinitiation or pr teolytic processing, the thermostable, enzymatically active, truncated forms of Taq Pol I directed by plasmids pFC84 and pFC85 (Table I) suggest that significant portions of the Taq Pol I sequence are not essential for DNA polymerase activity.

The purpose of the set of fusions of 5′ portions of the Taq Pol I A fragment with *lacZα* in BSM13⁺ was to confirm or determine the reading phase of the Taq Pol I gene internally as an aide to nucleotide sequencing. Since we knew the reading phase of *lacZ* in the BSM13⁺ polylinker, we could infer the reading phase of Taq Pol I in α-complementing in-frame fusions. DG98 harboring fusions which were in-frame were readily detectable as blue colonies on *X*-Gal indicator plates. We generated a series of fusions (Table II) at nine sites between nucleotides 962 and 1782 of the Taq Pol I gene.

We compared the DNA sequence of Taq Pol I with that of *E. coli* DNA polymerase I. At the DNA level, the two genes lack any significant regions of homology (Table III). In regions where the amino acid sequences are homologous, the DNA sequences diverge, especially in third positions of codons. The longest stretch of DNA sequence identity is 19 bases (Table III).

The predicted amino acid sequence of Taq Pol I is shown in Fig. 2. From this a codon bias table was generated (Table IV). There is a heavy bias toward G and C in the third position (91.8% C and G) as would be expected for GC-rich organisms

TABLE III

DNA sequence identity of Taq Pol I and E. coli Pol I

	Sequence location*	Nucleotide identity	Amino acid identity
Taq Pol I	190–208		
Pol I	178–196	19/19	6/6
Taq Pol I	1730–1757		
Pol I	2015–2042	23/28	9/9
Taq Pol I	2260–2277		
Pol I	2545–2562	17/18	6/6
Taq Pol I	2344–2363		
Pol I	2635–2654	17/20	7/7

* Nucleotide sequence coordinates for Taq Pol I from Fig. 2. Nucleotide sequence coordinates for *E. coli* Pol I adapted from GenBank.

TABLE IV

Codon usage in the T. aquaticus DNA polymerase I gene

Arg (76)	CGT	0	Leu (124)	TTG	3	Ser (31)	TCT	0
	CGC	24		TTA	0		TCC	15
	CGG	27		CTT	20		TCG	1
	CGA	0		CTC	46		TCA	0
	AGG	25		CTG	50		AGT	1
	AGA	0		CTA	5		AGC	14
Thr (30)	ACT	0	Pro (48)	CCT	3	Val (51)	GTT	1
	ACC	20		CCC	34		GTC	21
	ACG	10		CCG	9		GTG	29
	ACA	0		CCA	2		GTA	0
Ala (91)	GCT	2	Gly (58)	GGT	0	Ile (25)	ATT	3
	GCC	77		GGC	28		ATC	20
	GCG	12		GGG	30		ATA	2
	GCA	0		GGA	0			
Asn (12)	AAT	0	Gln (16)	CAG	15	Tyr (24)	TAT	4
	AAC	12		CAA	1		TAC	20
His (18)	CAT	0	Glu (87)	GAG	79	Cys (0)	TGT	0
	CAC	18		GAA	8		TGC	0
Asp (42)	GAT	3	Phe (27)	TTT	8	Lys (42)	AAG	37
	GAC	39		TTC	19		AAA	5
Met (16)	ATG	16	Trp (14)	TGG	14			

and as others have observed for other *Thermus* genes: 95% C and G for the *gk*24 gene encoding L-lactate dehydrogenase of *Thermus caldophilus* (15), 94.8% for *mdh* from *T. flavus* (14), and 89% for *leuB* from *T. thermophilus* (16).

Significant amino acid sequence similarity exists between Taq Pol I, *E. coli* Pol I, and bacteriophage T7 DNA polymerase. One possible sequence alignment yields 38% identity between the Taq Pol I and *E. coli* Pol I amino acid sequences (Fig. 7). There are two major regions of Taq Pol I and one region of T7 DNA polymerase that show extensive sequence similarity compared to *E. coli* Pol I. The first region of Taq Pol I extends from the N terminus to approximately residue 300. The second region extends from approximately residue 410 to the C terminus of Taq Pol I. The N-terminal region f Taq Pol I corresponds to the N-terminal domain of *E. coli* Pol I shown to contain the 5′-3′ exonuclease activity (17). The C-terminal regions of Taq Pol I and T7 DNA polymerase correspond to the *E. coli* Pol I domain shown to contain DNA polymerase activity (18). The x-ray structure of the Klenow fragment (19) shows that this domain contains a deep cleft believed to be responsible for DNA binding.

Apparently as a result of many mutations, deletions, insertions, etc. during evolution, Taq Pol I residues at positions 300–410 show little sequence similarity compared to *E. coli* Pol I. Taq Pol I is 96 residues shorter than *E. coli* Pol I; most of the deleted residues occur in the region encompassing residues 300–410. Ollis *et al.* (19) and Derbyshire *et al.* (20) have shown that *E. coli* Pol I residues Asp-355, Glu-357, Leu-361, Asp-424, Phe-473, and Asp-501 are involved in binding of divalent cation and deoxynucleoside monophosphate. A fragment of *E. coli* Pol I that contains only residues 515–928 is devoid of 3′-5′ exonuclease activity, but still retains polymerase activity (18). Presumably, the *E. coli* Pol I region comprised of residues 324–515 forms at least part, if not all, of the 3′-5′ exonuclease activity. Taq Pol I and *E. coli* Pol I display little sequence similarity in the presumptive 3′-5′ exonuclease region. Of the *E. coli* Pol I residues shown to be involved in cation and deoxynucleoside monophosphate binding, the sequence alignment of Fig. 7 shows only Asp-424 as having an exact homolog in the Taq Pol I sequence. Although other high scoring sequence alignments are possible in the Taq Pol I 300–410 region, it is possible that the Taq Pol I gene has undergone key mutations, deleti ns, or insertions

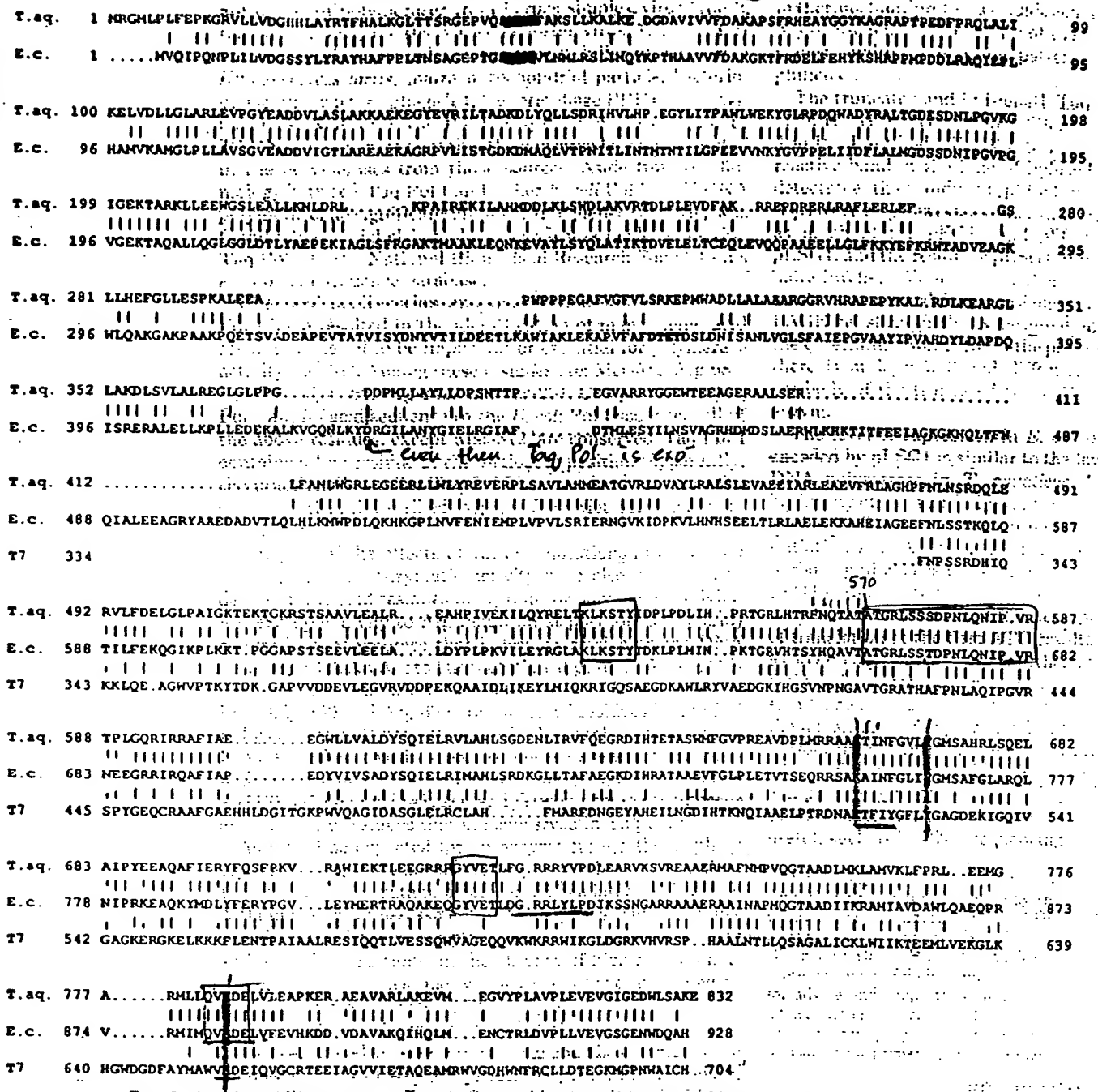


FIG. 7. Amino acid sequence comparison of the DNA polymerases from *T. aquaticus*, *E. coli*, and bacteriophage T7. Deduced amino acid sequences for Taq Pol I (*T.aq.*), *E. coli* Pol I (*E.c.*), and bacteriophage T7 DNA polymerase (*T7*) were analyzed for amino acid sequence homology by using the computer program GAP from the University of Wisconsin Genetics Computer Group. The alignment was obtained by using the mutational data scoring matrix of Staden (70). *Vertical marks* denote amino acid identities or functional relatedness between pairs of residues in the three sequences. *Half-vertical marks* denote amino acid identities or functional relatedness between residues in Taq Pol I and T7 DNA polymerase.

that have destroyed its 3′-5′ exonuclease activity. Preliminary results indicate that Taq Pol I displays little if any 3′-5′ exonuclease activity.

Sequence homology between *E. coli* Pol I and T7 DNA polymerase has been previously noted. Those T7 DNA polymerase sequences shown by Ollis *et al.* (21) to be conserved between that enzyme and *E. coli* Pol I are also present in the Taq Pol I amino acid sequence (Fig. 7). Most of the conserved residues are found in structural features that form the DNA-binding cleft of the enzyme. Although short segments of T7 DNA polymerase sequence in the 1–334 region are similar to regions in *E. coli* Pol I and Taq Pol I, the overall sequence similarity in this region, ignoring the first 300 residues of *E. coli* Pol I and Taq Pol I that form the 5′-3′ exonuclease domain, is poor. A complete and unambiguous sequence alignment for this region cannot be assigned. It should be noted that although T7 DNA polymerase also shows little similarity to *E. coli* Pol I in the region of the 3′-5′ exonuclease domain, T7 DNA polymerase is reported to display significant 3′-5′ exonuclease activity (22, 23).

Bernad *et al.* (24) and Pizzagalli *et al.* (25) have identified several short regions of DNA polymerase amino acid se-

quences that are highly conserved. These conserved sequences are found in polymerases from herpes simplex virus type 2, human cytomegalovirus, varicella-zoster virus, Epstein-Barr virus, vaccinia virus, adenovirus type 2, killer plasmid from *Kluveromyces lactis*, maize mitochondrial particle, bacteriophage ϕ29, bacteriophage T4, bacteriophage PRD1, and yeast plasmids. Neither *E. coli* Pol I, Taq Pol I, nor bacteriophage T7 DNA polymerase contains the conserved sequences noted in the polymerases from those sources. Aside from the homology between Taq Pol I and either *E. coli* Pol I or T7 DNA polymerase, no significant amino acid sequence similarity was found when a global homology search was made comparing Taq Pol I to the National Biomedical Research Foundation's amino acid sequence database.

Chemical modification and inactivation studies of *E. coli* Pol I have resulted in the identification of many amino acid residues believed to be important or essential for polymerase activity (26–31). Among these residues are: Met-512, Arg-682, Lys-758, Tyr-766, Arg-841, and His-881. Comparing the Taq Pol I amino acid sequence to the *E. coli* Pol I sequence, all of the above residues, except Met-512, are conserved. Taq Pol I contains a Leu residue at the analogous position. Apparently, the functionally similar Taq Pol I Leu residue at position 417 can fulfill the role ascribed to *E. coli* Pol I Met-512 in template primer binding (30).

Analyses of the effects of various mutations in the *E. coli* Pol I gene upon enzymatic activity have also been used to define amino acid residues important for polymerase activity. For example, a Gly to Arg mutation at position 850 (*polA5*) results in a polymerase that is less processive on the DNA substrate (32). An Arg to His mutation at position 690 (*polA6*) results in a polymerase that is defective in DNA binding (33). Both Gly-850 and Arg-690 are conserved residues in Taq Pol I. Joyce *et al.* (34) have characterized a number of *E. coli* Pol I mutants defective in 5′-3′ exonuclease activity. Interestingly, the four mutations, Y77C (*polA107*), G103E (*polA4113*), G184D (*polA480ex*), and G192D (*polA214*) all occur at amino acid residues that are conserved in Taq Pol I.

As would be expected for an enzyme from a thermophilic organism, Taq Pol I is considerably more thermostable than Pol I from *E. coli* (data to be presented in a later publication). Although a better assessment of an enzyme's thermostability would result from a complete cataloging of all stabilizing amino acid interactions, in the absence of high resolution x-ray crystal structures, many researchers have attempted to explain enzyme thermostability by an analysis of amino acid content (35–37). Several features of thermostable enzymes have been noted in such studies. Among those features are increased ratios of Arg to Lys residues, Glu to Asp residues, Ala to Gly residues, Thr to Ser residues, and a reduced Cys content. Comparing Taq Pol I to *E. coli* Pol I, the Ala to Gly and Thr to Ser ratios are smaller for Taq Pol I than for *E. coli* Pol I. Of the thermostabilizing type amino acid alterations that hold true, it is particularly notable that the Arg to Lys ratio for Taq Pol I is nearly twice that for *E. coli* Pol I. It is possible that the propensity of thermophilic proteins to contain Arg rather than Lys residues is simply a reflection of the high GC content of thermophilic organisms. The structural gene for Taq Pol I contains 67.9% GC compared to a 52.0% GC content for *E. coli* Pol I. The six Arg codons are rich in G and C (13 out of 18 bases are G or C) compared to the two Lys codons (1 out of 6 bases is a G). This explanation for amino acid preferences in proteins from thermophilic organisms cannot be the basis for Glu *versus* Asp, Thr *versus* Ser, or Ala *versus* Gly preference, because there are equal ratios of GC *versus* AT in the codons for those pairs of amino acids. A more likely explanation for the preference for Arg over Lys in thermostable proteins would seem to be based on the unique physical-chemical properties of the two amino acids (*e.g.* pK_a values, hydrogen bonding patterns, hydrophobicity/hydrophilicity).

The truncated and full-length Taq Pol I enzymes produced upon IPTG induction show different reactivities to the anti-Taq Pol I antibody. For Western blots (Fig. 5), the immunoreactive band in the lane of induced pLSG1 is more readily detectable than induced pFC84 or pFC85, the *Sph*I and *Asp*718 A fragment deletions. In fact, we loaded three times as much of the pFC84 and pFC85 extracts compared to pLSG1, and the resulting pLSG1 immunoreactive band is still more intense. We infer that there are more epitopes for our antibody, prepared from full-length (94-kDa) Taq Pol I SDS-PAGE gel slice, in the N-terminal end of Taq Pol I than in the C-terminal two thirds of the protein. Or, based on activity, there is at least a 3-fold difference in reactivity with the antibody of the truncated *versus* the full-length form of the enzyme.

The level of expression in *E. coli* of full-length Taq Pol I encoded by pLSG1 is similar to the level of expression of *Taq* DNA polymerase in *T. aquaticus*. In pLSG1 (Fig. 6) the beginning of the Taq Pol I open reading frame is 109 bp distal to the *Bgl*II site and 171 bp distal to the *lacZ*α translation initiation site. A low level of Taq Pol I expression in cells harboring pLSG1 is consistent with an in-phase TGA codon (−111 through −109, Fig. 2) in the *Taq* DNA sequence causing translation termination of the *lacZ*α polypeptide. Reinitiation of translation at the first ATG results in the synthesis of the 94-kDa Taq Pol I protein. Further manipulation of the *Taq* DNA polymerase sequence has increased the level of expression.[4] The cloned full-length Taq Pol I gene in pLSG1 affords the advantages of expressing Taq Pol I in *E. coli* and in ease of isolating the enzyme from *E. coli* compared to *T. aquaticus*. These advantages will aid in further study of the enzyme and will provide a ready source of Taq Pol I for use in PCR and other biochemical procedures in which Taq Pol I might prove useful, such as in DNA sequencing.

Acknowledgments—We gratefully acknowledge Gail Rodgers for advice on preparation of antibody from a small amount of protein; Corey Levenson, Lauri Goda, and Dragan Spasic for preparation of oligonucleotide primers, Keith Bauer for fermentation support; Will Bloch and David Birch for advice on nonradioactive detection of proteins; Henry Erlich, Tom White, John Sninsky, and Michael Innis for advice and support; Hamilton Smith and Norman Arnheim for advice and critical review of the manuscript; Sharon Nilson and Eric Ladner for preparation of figures; and Patricia A. Robinson and Edna McCallan for preparation of the manuscript.

REFERENCES

1. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985) *Science* 230, 1350–1354
2. Mullis, K. B., and Faloona, F. A. (1987) *Methods Enzymol.* 155, 335–350
3. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* 239, 487–491
4. Chien, A., Edgar, D. B., and Trela, J. M. (1976) *J. Bacteriol.* 127, 1550–1557
5. Kaledin, A. S., Slyusarenko, A. G., and Gorodetskii, S. I. (1980) *Biokhimiya* 45, 644–651
6. Raleigh, E. A., Murray, N. E., Revel, H., Blumenthal, R. M., Westaway, D., Keith, A. D., Rigby, P. W. J., Elhai, J., and Hanahan, D. (1988) *Nucleic Acids Res.* 16, 1563–1575
7. Murizna, N. V., Vorozheykina, D. P., and Matvienko, N. I. (1988) *Nucleic Acids Res.* 16: 8172

[4] F. C. Lawyer, S. Stoffel, and D. H. Gelfand, unpublished observations.

8. Iijima, S., Uozumi, T., and Beppu, T. (1986) *Agric. Biol. Chem.* **50**, 589–592
9. Nagahari, K., Koshikawa, T., and Sakaguchi, K. (1980) *Gene (Amst.)* **10**, 137–145
10. Slatko, B. E., Benner, J. S., Jager-Quinton, I., Moran, L. S., Simcox, T. G., Vann Cott, E. M., and Wilson, G. G. (1987) *Nucleic Acids Res.* **15**, 9781–9796
11. Kelley, W. S., Chalmers, K., and Murray, N. E. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5632–5636
12. Lin, T. C., Rush, J., Spicer, E. K., and Konigsberg, W. H. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 7000–7004
13. O'Farrell, P. H., Polisky, B., and Gelfand, D. H. (1978) *J. Bacteriol.* **134**, 645–654
14. Nishiyama, M., Matsubara, N., Yamamoto, K., Iijima, S., Uozumi, T., and Beppu, T. (1986) *J. Biol. Chem.* **261**, 14178–14183
15. Kunai, K., Machida, M., Matsuzawa, H., and Ohta, T. (1986) *Eur. J. Biochem.* **160**, 433–440
16. Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T., and Oshima, T. (1984) *J. Biol. Chem.* **259**, 2956–2960
17. Brutlag, D., Atkinson, M. R., Setlow, P., and Kornberg, A. (1969) *Biochem. Biophys. Res. Commun.* **37**, 982–989
18. Freemont, P. S., Ollis, D. L., Seitz, T. A., and Joyce, C. M. (1986) *Proteins Struct. Funct. Genet.* **1**, 66–73
19. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985) *Nature* **313**, 762–766
20. Derbyshire, V., Freemont, P. S., Sanderson, M. R., Beese, L., Friedman, J. M., Joyce, C. M., and Steitz, T. A. (1988) *Science* **240**, 199–201
21. Ollis, D. L., Kline, C., and Steitz, T. A. (1985) *Nature* **313**, 818–819
22. Tabor, S., and Richardson, C. C. (1987) *J. Biol. Chem.* **262**, 15330–15333
23. Tabor, S., Huber, H. E., and Richardson, C. C. (1987) *J. Biol. Chem.* **262**, 16212–16223
24. Bernad, A., Zaballos, A., Salas, M., and Blanco, L. (1987) *EMBO J.* **6**, 4219–4225
25. Pizzagalli, A., Valsasnini, P., Plevani, P., and Lucchini, G. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 3772–3776
26. Mohan, P. M., Basu, A., Basu, S., Abraham, K. I., and Modak, M. J. (1988) *Biochemistry* **27**, 226–233
27. Basu, A., and Modak, M. J. (1987) *Biochemistry* **26**, 1704–1709
28. Pandey, V. N., Williams, K. R., Stone, K. L., and Modak, M. J. (1987) *Biochemistry* **26**, 7744–7748
29. Pandey, V. N., and Modak, M. J. (1988) *J. Biol. Chem.* **263**, 6068–6073
30. Basu, A., Williams, K. R., and Modak, M. J. (1987) *J. Biol. Chem.* **262**, 9601–9607
31. Joyce, C. M., Ollis, D. L., Rush, J., Steitz, T. A., Konigsberg, W. H., and Grindley, N. D. F. (1986) *UCLA Symp. Mol. Cell. Biol. New Ser.* **32**, 197–205
32. Matson, S. W., Capaldo-Kimball, F. N., and Bambara, R. A. (1968) *J. Biol. Chem.* **253**, 7851–7856
33. Kelly, W. S., and Grindley, N. D. F. (1976) *Nucleic Acids Res.* **3**, 2971–2984
34. Joyce, C. M., Fujii, D. M., Laks, H. S., Hughes, C. M., and Grindley, N. D. F. (1985) *J. Mol. Biol.* **186**, 283–293
35. Singleton, R., and Amelunxen, R. E. (1973) *Bacteriol. Rev.* **37**, 320–342
36. Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frand, G., and Tratschin, J. D. (1979) *Biochemistry* **18**, 5698–5703
37. Ponnuswamy, P. K., Muthusamy, R., and Manavalan, P. (1982) *Int. J. Biol. Macromol.* **4**, 186–190
38. Young, R. A., and Davis, R. W. (1983) *Science* **222**, 778–782
39. Wood, W. B. (1966) *J. Mol. Biol.* **16**, 118–133
40. Casadaban, M. J., and Cohen, S. N. (1980) *J. Mol. Biol.* **138**, 179–207
41. Cole, G. E., McCabe, P. C., Inlow, D., Gelfand, D. H., Ben-Bassat, A., and Innis, M. A. (1988) *BioTechnology* **6**, 417–421
42. Miller, J. H. (1972) *Experiments in Molecular Genetics*, pp. 201–205, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
43. Marinus, M. G., Carraway, M., Frey, A. Z., Brown, L., and Arraj, J. A. (1983) *Mol. Gen. Genet.* **192**, 288–289
44. Young, R. A., and Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 1194–1198
45. Loenen, W. A. M., and Blattner, F. R. (1983) *Gene (Amst.)* **26**, 171–179
46. Russel, M., Kidd, S., and Kelley, M. R. (1986) *Gene (Amst.)* **45**, 333–338
47. Greene, P. J., Betlach, M. C., Goodman, H. M., and Boyer, H. (1974) *Methods Mol. Biol.* **7**, 87–111
48. Modrich, P., and Zabel, D. (1976) *J. Biol. Chem.* **251**, 5866–5874
49. Bingham, A. H. A., Sharman, A. F., and Atkinson, T. (1977) *FEBS Lett.* **76**, 250–256
50. Bickle, T. A., Pirrotta, V., and Imber, R. (1977) *Nucleic Acids Res.* **4**, 2561–2572
51. Gingeras, T. R., Myers, P. A., Olson, J. A., Hanberg, F. A., and Roberts, R. J. (1978) *J. Mol. Biol.* **118**, 113–122
52. Moore, S. K., and James, E. (1976) *Anal. Biochem.* **75**, 545–554
53. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580
54. Birnboim, H. C., and Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523
55. Maniatis, T., Fritch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
56. Castenholz, R. W. (1969) *Bacteriol. Rev.* **33**, 476–504
57. Young, R. A., and Davis, R. W. (1985) *Genetic Engineering* (Setlow, J. K., and Hollaender, A., eds) Vol. 7, pp. 29–41, Plenum Publishing Corp., New York
58. Sheldon, E. L., Kellogg, D. E., Watson, R., Levenson, C. H., and Erlich, H. A. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 9085–9089
59. Huynh, T. V., Young, R. A., and Davis, R. W. (1985) in *DNA Cloning, IRL Press, Oxford Techniques: A Practical Approach* (Glover, D., ed) pp. 50–77, IRL Press, Oxford
60. Wong, G., Arnheim, N., Clark, R., McCabe, P., Innis, M., Aldwin, L., Nitecki, D., and McCormick, F. (1986) *Cancer Res.* **46**, 6029–6033
61. Mack, D. H., Bloch, W., Nath, N., and Sninsky, J. J. (1988) *J. Virol.* **62**, 4786–4790
62. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., and Dreyfuss, G. (1986) *Mol. Cell. Biol.* **6**, 2932–2943
63. Goldstein, L. S. B., Laymon, R. A., and McIntosh, J. R. (1986) *J. Cell Biol.* **102**, 2076–2087
64. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78
65. Sanger, F. (1981) *Science* **214**, 1205–1210
66. Barnes, W. M., Bevan, M., and Son, P. H. (1983) *Methods Enzymol.* **101**, 98–122
67. Gough, J. A., and Murray, N. E. (1983) *J. Mol. Biol.* **166**, 1–19
68. Mizusawa, S., Nishimura, S., and Seela, F. (1986) *Nucleic Acids Res.* **14**, 1319–1324
69. Innis, M. A., Myambo, K. B., Gelfand, D. H., and Brow, M. A. D. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9436–9440
70. Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951–2961

Continued on next page.

SUPPLEMENTARY MATERIAL

Isolation, Characterization, and Expression in E. coli of the DNA Polymerase Gene from Thermus aquaticus

Frances C. Lawyer, Susanne Stoffel, Randall K. Saiki, Kenneth Myambo, Robert Drummond, and David H. Gelfand

MATERIALS AND METHODS

Bacterial strains – Thermus aquaticus strain YT1, ATCC 25104, was obtained from The American Type Culture Collection. E. coli strains Y1088 [supE supF metB trpR hsdR⁻ hsdM⁺ tonA21 strA lacU169 proC::Tn5 (pMC9)] and Y1090 [lacU169 proA⁺ Δlon araD139 strA supF trpC Tn10 (pMC9)] (38) were a gift from R.A. Young and were used for plating the λgt11 libraries and in making lysogens of candidate clones. E. coli strains K802 (hsdR⁻ hsdM⁺ mcrB⁻ gal⁻ metB supE) (39, 4[illegible]) and MC1000 (araD139 Δ(ara,leu)7697 ΔlacX74 galU galK strA) (40) were used to plate the λCh35 library. The derivation of E. coli strain DG98 (thi-1 endA1 hsdR17 lacI^q lacZΔM15 proC::Tn10 supE44 /F' lacI^q lacZΔM15 proC⁺) has been described (41). Strain DG101 (thi-1 endA1 hsdR17 lacI^q lacZΔM15 supE44) is a proC⁺ female derivative of strain DG98 prepared in this laboratory. Strain DG105 (thi-1 endA1 hsdR17 lacI^q lacZΔM15 supE44 dam13::Tn9 F'lacI^q lacZΔM15) is a dam⁻ derivative of DG98 prepared in this laboratory by using bacteriophage P1 transduction (42) and E. coli K-12 strain GM1894 (43) as the source of the dam13::Tn9 insertion.

Bacteriophages and Plasmids – EcoRI-digested, alkaline phosphatase-treated λgt11 (44) was purchased from Promega Biotech. λCh35 (45) was provided by F. Blattner. R408 helper phage (46) was used to generate single-stranded DNA from BSM13⁺ subclones. Plasmid pTZ19R was purchased from United States Biochemical Corp. (USB), and plasmid BSM13⁺ was purchased from Stratagene.

Cloning Reagents – Most restriction enzymes were purchased from New England BioLabs (NEB). Asp718 was purchased from Boehringer Mannheim. EcoRI and XhoI were prepared in house (47, 48, 49, 50, 51). T4 DNA ligase was also prepared in house (52). Polynucleotide kinase was purchased from NEB and was used essentially according to manufacturer's instructions for kinasing oligonucleotide linkers. E. coli DNA polymerase I, large fragment (Klenow) was purchased from NEB and was used according to manufacturer's instructions for filling in 5'-sticky-ends generated by restriction digests.

8-mer EcoRI linkers were purchased from Collaborative Research. 10-mer and 12-mer EcoRI linkers and 8-mer BglII linkers were purchased from NEB.

[α-^{32}P]-labeled dNTP's, 800 Ci/mmol, and [γ-^{32}P]-ATP, 3000 Ci/mmol were purchased from New England Nuclear.

Other Cloning Procedures – Plasmids were transformed into DG98, DG101, and DG105 as described (53). Plasmid mini DNA preparations and λ phage mini DNA preparations were carried out as described (54, 55).

Immunological Reagents – Rabbit polyclonal preimmune serum and high-titer Taq Pol I (94-kDa) immune serum were prepared by Berkeley Antibody Company (Bab Co). Rabbits were immunized intranodally with 20 μg of homogeneous 94-kDa Taq Pol I in a homogenized SDS-PAGE gel slice (in complete Freund's adjuvant). Rabbits were boosted intranodally at 3-week intervals with 10 μg of immunogen in incomplete Freund's adjuvant. For screening λgt11:Taq libraries and for Western blots, antibody was preabsorbed for at least 2 h with 200 μg E. coli crude extract per 6 μl undiluted serum in 10X phosphate buffered saline (PBS) and 1% Tween 20.

Goat anti-rabbit IgG horseradish peroxidase conjugate was purchased from Bio Rad.

Isolation of DNA from T. aquaticus - Taq strain YT1 was grown in medium D with 0.1% tryptone and 0.1% yeast extract (56) in a 14 liter Chemap fermentor at air saturation, 70 °C and 400 rpm. Cells were pelleted and frozen at -20 °C. A sample (2 g) of frozen cell paste was suspended in 10 ml 25% sucrose, 50 mM Tris-HCl, pH 8.0, and 0.25 M EDTA pH 8.0. Lysozyme (Sigma) was added to 250 μg/ml. After incubation at 37 °C for 10 min, 200 ml pretreated (37 °C for 20 min) proteinase K solution (10 mM Tris-HCl, pH 8.0, 10 mM EDTA, 10 mM NaCl, 0.5% SDS, 50 μg/ml proteinase K (Sigma)) was added. The mixture was incubated overnight at 37 °C. The mixture was extracted with an equal volume of phenol saturated with TE. Aqueous and phenol phases were reextracted, and the aqueous phases pooled and extracted with an equal volume of 24:1 chloroform:isoamyl alcohol. The extracted material was dialyzed against many changes of TE plus 10 mM NaCl over the course of several days. The dialysate was then treated with 50 μg/ml RNase A and 10 units/ml RNase T1 (Calbiochem), incubating 4 h at 37 °C. SDS was added to 0.5% and Proteinase K to 50 μg/ml, followed by extractions with TE-saturated phenol and chloroform-isoamyl alcohol. Sodium acetate was added to 0.3 M, 2 volumes of 95% ethanol were layered over the aqueous layer, and the DNA was gently spun out. The DNA was rinsed in 70% ethanol and then 95% ethanol and was rehydrated in TE over several days.

Preparation and screening of λgt11 libraries – Taq DNA was digested to completion with AluI, yielding DNA fragments in the range of 0.1 to 1 kb. The AluI fragments were ligated in three separate reactions to [γ-^{32}P]-ATP kinased, annealed 8-mer, 10-mer, and 12-mer EcoRI linkers and excess linkers were removed as described (59). EcoRI-digested, dephosphorylated λgt11 vector was ligated with EcoRI-digested targets. In vitro packaging of the ligations was performed with Gigapack Plus packaging components (Stratagene), using conditions suggested by the supplier. Packaging mixtures were plated on Y1090 for direct antibody screening or to make plate lysates for subsequent antibody screening.

The libraries were immunoscreened essentially as described (57). Briefly, plates containing between 5 x 10^4 and 1 x 10^5 plaques on lawns of Y1090 were incubated for 3.5 h at 42 °C. The plates were then overlaid with nitrocellulose filters (Schleicher & Schuell) previously soaked in 10 mM IPTG and were incubated at 37 °C for 2 h. Filters were removed from the plates and blocked overnight at 4 °C in a solution of 1 M glycine, 5% non-fat dry milk and 1% ovalbumin. After blocking, filters were washed (PBS, pH 6.8, 0.1% Tween 20, 0.1% nonfat dry milk) and then were incubated with a 1:6000 dilution of pretreated primary antibody in wash solution at room temperature for 3 h. After washing, filters were incubated 1 h at room temperature with a 1:3000 dilution of goat-anti-rabbit:HRP in wash solution. After further washing, filters were fixed in 1% dextran sulfate, 10 mM sodium citrate buffer and 10 mM EDTA, pH 5.0, for 15 min. The color was developed in 10 mM sodium citrate, 10 mM EDTA, pH 5.0, buffer plus 0.4 mM TMB and 45 μl 3% H_2O_2 per 100 ml reaction buffer (58). Positive plaques were plaque purified, rescreened with antibody, and saved for further study.

Lysogens of λgt11 clones in E. coli strain Y1089 were prepared and induced to make fusion proteins for Western blotting as described (59).

Preparation and Screening of a λCH35:Taq library – Taq genomic DNA was subjected to partial digest with Sau3A, run on 10-40% sucrose gradients (55), and fractions containing Sau3A fragments in the range of 10-20 kb were pooled, dialyzed against several changes of TE and ethanol precipitated. λCh35 cos ends were ligated at 37 °C for 1 h. Following heat inactivation of the ligase, the vector was digested with BamHI. Arms were purified by pelleting through 5-20% potassium acetate gradients (55). Target and vector were ligated, ligations were packaged with Gigapack Plus extracts (Stratagene) according to the manufacturer's instructions, and the packaging reactions were plated on host K802 to make plate lysates. Plate lysates were prepared as described (55) yielding a library which had arisen from approximately 5 x 10^5 individual plaques.

The library was screened (55), plating 10^4 plaques per plate on host MC1000 (to eliminate lacI-Z hybridization background). The library lifts were probed (55), by using the alkali-denatured probe described below. Positive plaques from the probing of the library were subjected to two rounds of plaque purification and probe screening to obtain pure positive plaques for further study.

LacZα fusions – To confirm the frame at the NheI site (nucleotide 2043, Fig. 2), pFC85 was digested with NheI and KpnI and treated with Klenow and all four dNTPs. The Klenow-treated mixture was diluted to 5 μg/ml, ligated, transformed into DG98, and the transformation mixture was plated on X-gal plates. We determined the DNA sequence at the fusion site of one blue candidate which had the NheI site regenerated. From this we deduced the reading phase of the Taq Pol I sequence 5' of the NheI site (Table II).

Fusions in two different reading frames were constructed at the BamHI site (nucleotide 1780, Fig. 2). One was made by digesting pFC85 with BamHI, diluting the digest, and ligating. The resultant colonies, upon transformation into DG98, were all white on X-gal plates; thus the fusion was out of frame. One candidate of the appropriate size was sequenced at the fusion site to confirm that the sequence was as expected (Table II). The other fusion at the BamHI site was made by digesting pFC85 with BamHI, repairing the 5' sticky end with Klenow and all four dNTPs, diluting, and ligating. The DNA sequences of two transformant candidates with different properties were determined. Candidate #33 was blue and had a ClaI site (when cells were propagated in dam⁻ strain DG105) and no BamHI site as expected for this particular fusion. Candidate #35 was also blue but had a BamHI site and no ClaI site. This clone had sustained a three base pair deletion which regenerated the BamHI site while maintaining the frame with respect to lacZ.

A family of XhoI (nucle[illegible] Fig. 2) deletions was constructed by using XhoI enzyme containing a 3'-5' exonuclease a[illegible]-C35 was digested with XhoI and either Asp718 or SmaI, the ends were repaired with Klenow and all four dNTP's, and the mixture was diluted and ligated. The DNA sequence at the Taq:lacZ fusion site was determined for several different DG98 transformants, both blue and white, having sustained varying sized deletions. The deletions ranged from 1-bp within the XhoI site (#28) to 450-bp to the left of the XhoI site (#30). See Table II for a summary of the deletions and their properties.

Assembly of the full-length Taq Pol I gene – To facilitate assembly of the Taq Pol I gene, vector BSM 13⁺ was modified so that it would have a BglII site at the position of the HindIII site in the polylinker and maintain the same lacZα reading frame as that of BSM13⁺. HindIII-digested, Klenow repaired BSM13⁺ was ligated with kinased and annealed 8-mer BglII linker. Linker plus vector templates were gel purified, digested with BglII, ligated, diluted, and transformed into DG98. Blue candidates on X-gal plates were screened and a correct candidate (identified by DNA sequence analysis) with a BglII site and without a HindIII site was chosen to be the Taq Pol I cloning vector.

The BSM13⁺ BglII vector was digested with BglII and Asp718 and ligated in a three fragment ligation with gel purified target fragments: the 724-bp BglII-HindIII portion of the B fragment and the 2.8-kb HindIII - Asp718 portion of the A fragment. Transformants in DG98 were screened by alkaline-SDS mini screens, and one correct candidate was designated pLSG1 (Fig. 4).

Probes – For probing Taq genomic Southern blots, the Taq sequence in clone λgt11:[illegible] was amplified via polymerase chain reaction (PCR) as described (1), adding 150 μCi [α-^{32}P]-dCTP to the reaction.

Probe for screening the λCh35 library was prepared by alkali denaturation and Klenow extension of pTZ19R:[illegible]. 10 μg of pTZ19R:[illegible] was denatured at room temperature for 5 min in [illegible] μl containing 0.2 M NaOH and 2 mM EDTA. Ammonium acetate, pH 4.5, was added to 0.2 M and the mixture was ethanol precipitated, washed, and dried in a vacuum concentrator (Savant Speed Vac). The denatured plasmid was annealed with 10 pmol universal sequencing primer in a [illegible] μl reaction mixture containing 20 mM Tris-HCl, pH 8.0, 10 mM MgCl$_2$, and 50 mM KCl at 37 °C for 15 min. To 10 μl of the mixture was added 500 pmol each dATP and TTP, 200 pmol each dGTP and dCTP, 4 μl each of [α-^{32}P]-dCTP and [α-^{32}P]-dGTP, KCl to 0.1 M and 5 units Klenow, in a final volume of 30 μl. The extension reaction was incubated at 37 °C for 15 min and then stored at -20 °C.

Southern Blots – Southern blots were performed as described (55) with the following exceptions: DNA was transferred onto MSI Magnagraph nylon membrane instead of nitrocellulose. Prehybridization solution was 50% formamide, 5X SSPE, 0.1% SDS, 5X Denhardt's solution and 200 μg/ml sheared salmon sperm DNA. Hybridization solution was the same except for 2X Denhardt's solution, 100 μg/ml denatured sheared salmon sperm DNA and 10^6 cpm [α-^{32}P]-labeled PCR-amplified probe. Prehybridization and hybridization were both carried out at 42 °C.

Western Blots – Western blots and nonradioactive detection of bound antigen were carried out as described (60, 61), omitting FBS from the blocking and washing solutions, and substituting sodium citrate for sodium fumarate in the substrate solution (58).

Epitope Selection – Liquid cultures of λgt11:Taq lysogens were induced as follows: overnight cultures in L broth plus 0.2% glucose were diluted to OD_{600} = 0.04 and grown in the same medium at 30 °C to OD_{600} = 0.5. Cultures were heat induced at 40 °C for 15 min, then IPTG was added to 10 mM and the cultures were further incubated at 37 °C for 1 h. For epitope selection (62, 63), cultures were pelleted and pellets were suspended in 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1% SDS before sonicating 3 min. After sonication, β-mercaptoethanol was added to 5% and glycerol to 10%; samples were loaded on SDS-PAGE gels and the gels were run at 40 V overnight. Proteins were electroblotted onto nitrocellulose filters. The filters were stained with Ponceau S to localize the fusion proteins, then blocked and washed as above. Fusion protein bands were excised and incubated with 1.5 ml of a 1:200 dilution of anti-Taq Pol I antibody in TBST (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.05% Tween 20) overnight at 4 °C. Absorbed antibody strips were sequentially washed with PBS and TBST, and then transferred to syringes. After further TBST washes, the absorbed antibody was eluted from the strips with 1 ml 0.2 M glycine, pH 2.8, and expelled into 1 ml 100 mM Tris base. An aliquot of 0.2 ml of the eluted antibody was used to probe Western blots containing lanes of both purified Taq Pol I and T. aquaticus crude extract (Fig. 1).

IPTG Inductions – Early log phase cultures of BSM13⁺ derivatives were induced with 2.5 - 10 mM IPTG, grown to late log phase (OD_{600} = 1), and pelleted. Pellets were stored frozen at -70 °C. For SDS-PAGE and for Taq Pol I activity assays, thawed pellets were resuspended in 1 ml sonication buffer (50 mM Tris-HCl, pH 7.4, 1 mM EDTA, 2.5 mM PMSF, 0.5 μg/ml leupeptin) and sonicated. For SDS-PAGE, the appropriate amount of sonicate was added to loading buffer (1 X = 50 mM Tris-HCl, pH 6.8, 10% glycerol, 1% β-mercaptoethanol, 1% SDS, 2 mM EDTA, 0.05% bromophenol blue) and 20 μg carrier protein. For Taq Pol I assays, the appropriate amount of sonicate was diluted in assay buffer and heat-treated as described below.

Preparation of Extracts and Enzyme Activity Assays – Induced or uninduced cell pellets were prepared as described above. Extracts were prepared by resuspension of the cell pellets at 10 OD_{600}/ml in 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 0.5 μg/ml leupeptin (Sigma) and 2.4 mM PMSF (diluted from a 144 mM stock in dimethyl formamide) and sonication (6 min, 50% duty cycle, [illegible] output, Branson cup sonifier with circulating water at 0 °C). The lysed cell suspension was diluted with an equal volume of 10 mM Tris-HCl, pH 8, 50 mM KCl, 0.5% NP-40, 0.5% Tween 20, 1 mM EDTA and heated at 75 °C for 20 min to inactivate and denature E. coli host proteins. The heated extracts were chilled at 0 °C for 15 min and centrifuged at 12,000 X G for 10 min. The clarified extracts were diluted as necessary in 25 mM Tris-HCl, pH 8 (20 °C), 50 mM KCl, 1 mg/ml autoclaved gelatin, 1 mM β-mercaptoethanol, 0.5% NP-40, 0.5% Tween 20.

Taq Pol I activity was assayed at 74 °C for 10 min in a 50 μl reaction containing 25 mM N-tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid (TAPS) pH 9.5 (20 °C), 50 mM KCl, 2 mM MgCl$_2$, 200 μM each TTP, dATP, and dGTP, 100 μM [α-^{32}P]-dCTP (0.05 μCi/nMole), 12.5 μg activated salmon sperm DNA, 1 mM β-mercaptoethanol and 5 μl diluted enzyme. Reactions were stopped by the addition of 10 μl 60 mM EDTA and stored at 0 °C. Aliquots (50 μl) were diluted with 1 ml of 2 mM EDTA containing 50 μg/ml sheared salmon sperm DNA, and precipitated by the addition of 1 ml 20% trichloroacetic acid (w/v) and 2% sodium pyrophosphate, and incubation at 0 °C for 15 min. Precipitated DNA was collected on GF/C filter discs (2.4 cm) and washed extensively with 5% trichloroacetic acid and 2% sodium pyrophosphate (35 ml), then 95% ethanol (5 ml), dried, and counted. One unit corresponds to 10 nmoles of product synthesized in 30 min. DNA polymerase activity is linearly proportional to enzyme concentration up to 80-100 pmoles dCMP incorporated (diluted enzyme at 24-30 munits/μl).

For the first experiment in Table I, total dCTP was reduced to 50 μM, [α-^{32}P]-dCTP was 0.25 μCi/nMole, and incubation time was 20 min.

DNA Sequencing – The DNA sequencing strategy is detailed in Fig. 4. Many of the subclones described above in both Results and Materials and Methods were used in deriving the DNA sequence of the Taq Pol I gene: pFC43 and the BglII deletion of the B fragment, pFC82.35, pFC84, pFC85 and the lacZ α fusion plasmids. In addition, several plasmids were constructed for deriving further DNA sequences from the A fragment. To obtain sequence rightward of the NheI site, pFC85 was digested with HindIII and NheI, and the 5' sticky ends repaired with Klenow. The mixture was diluted, ligated, and transformed into DG98. A correct candidate was identified by restriction mapping. To obtain DNA sequence rightward of the BamHI site in the 8.0-kb HindIII A fragment, pFC85 was digested with HindIII and BamHI, repaired with Klenow and all four dNTP's, diluted, ligated, and transformed into DG98. Upon miniscreening, a candidate was chosen which was the appropriate size and which had the expected PvuII digest pattern. Several restriction fragments were also cloned into M13mp10 and M13mp11 phage vectors (64): the 1418-bp SacI-HindIII portion of the B fragment and the 1163-bp HindIII-BamHI and 1517-bp BamHI-Asp718 portions of the A fragment (Fig. 4).

In addition to the universal sequencing primer and the reverse sequencing primer, several primers were synthesized based on previously generated sequence. Primers are listed in Table V. DNA sequence analysis was performed by the dideoxy chain termination method (65) using M13 subclones or single-stranded templates generated from BSM13⁺ Taq clones. Due to the high GC content of the gene, recurrent band compressions and occasional false terminations as a result of strong secondary structure were observed on the sequencing gels. The compressions were resolved by substituting deaza-7-dGTP or inosine (dITP) (66, 67, 68) for dGTP in the normal reactions. The DNA sequence for several regions was confirmed by sequencing at 70 °C with Taq Pol I (69).

Single-stranded DNA was generated from BSM13⁺ derivatives by using R408 helper phage by a variation of the method of R. Russell et al. (46). Prewarmed L broth containing 0.1% glucose and 100 μg/ml Ampicillin was inoculated with a single colony of the desired BSM13⁺ derivative. The culture was grown at 37 °C to an OD_{600} = 0.2 and then inoculated with R408 at multiplicity of infection of approximately 5 to 10. The infected culture was aerated vigorously at 37 °C for 4 h, then single stranded DNA was isolated as described (46).

Table V
DNA Sequencing Primers

Primer	Taq Pol I Nucleotide Coordinates[a]
MK122	878 → 897
MK123	859 → 876
MK124	841 → 879
MK130	2098 ← 2127
MK131	1873 ← 1891
MK132	1651 ← 1668
MK133	1730 ← 1749
MK134	1588 ← 1606
MK135	291 ← 309
MK136	295 ← 311
MK137	355 ← 373
MK139	1013 ← 1032
MK140	136 ← 156
MK141	44 ← 62
MK142	1015 ← 1034
MK143	1313 ← 1332
MK144	1115 ← 1135
MK145	2140 → 2159
MK148	2140 → 2159
MK150	358 → 377
MK151	2561 → 2580
MK155	297 → 278
MK158	130 → 144
MK159	64 → 48

[a] Arrows denote strand sequenced with each primer: → in sense strand, 5′ to 3′, and ← in antisense strand, 3′ to 5′.

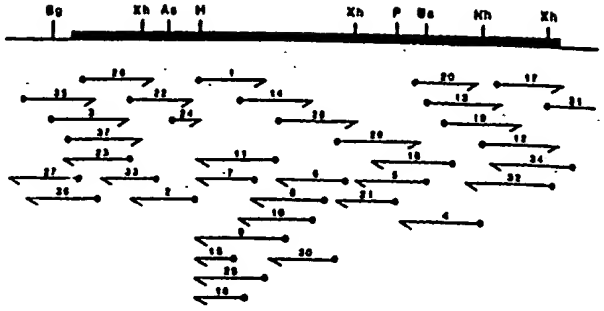


Fig. 8. DNA sequencing strategy. The bold line on the map delineates the coding sequencing for Taq Pol I. Arrows indicate sequence obtained in the sense (→) or non-sense (←) direction. Length of the arrow corresponds to the amount of sequence obtained in each case. 100% of the DNA sequence was determined on both strands. 1-3, the universal sequencing primer was used for sequencing on templates pFC82.35, pFC83, and the B fragment BglII-HindIII subclone. 4-11, the universal sequencing primer was used with Exo7m deletion templates ΔNhe #1, ΔBam #15, ΔXho #28, #30, #32, #33, #34, and #50. 12-13, reverse sequencing primer was used with A-fragment deletions MluI-SstI and MluI-BamHI. 14-37, primers utilized were MK122-124, MK130-136, MK138-143, MK148, MK150, MK151, MK155, MK158, and MK159.